Bayesian Detection and Estimation of Noisy Sinusoids with Reversible Jump Simulated Annealing

DURSUN ÜSTÜNDAĞ[1], MEHMET CEVRİ[2]
[1]Department of Mathematics, Faculty of Science,
Marmara University,
Istanbul,
TURKEY

[2]Department of Mathematics, Faculty of Science,
Istanbul University,
Istanbul,
TURKEY

Abstract: - This paper presents a novel algorithm that combines a global nonlinear optimization routine based on the Bayesian Inference Reversible Jump Markov Chain Monte Carlo (BI-RJMCMC) method under various proposal distributions with Simulated Annealing (SA). It performs a global search in the joint space of the parameters and number of parameters, thereby surmounting the problem of local minima and converging to the modes of the full posterior distribution efficiently. Finally, the algorithm is used for detecting the number of sinusoids and estimating their parameters from corrupted data. The results of all the simulations support the effectiveness and reliability of the algorithm.

Key-Words: - Signal Processing, Model Selection, Parameter Estimation, Reversible Jump, Markov Chains Monte-Carlo Sampling, Bayesian Inference, Simulated Annealing.

Received: April 2, 2024. Revised: October 14, 2024. Accepted: November 17, 2024. Published: December 31, 2024.

1 Introduction

The ability to comprehend and anticipate the outcomes of mathematical modeling is a crucial skill in the fields of science and engineering. In the case of real-world phenomena, a model, which approximates some extent, is used to describe a physical process that generates a specific signal or set of observations. Ultimately, it is pertinent to inquire which model most accurately represents the data at hand and to identify the optimal estimate of the free parameters that characterize the model of interest. Consequently, a significant challenge arises in a multitude of applications, namely the extraction of these distinctive characteristics from physical processes based on observations.

Given the extensive applicability of sinusoidal models, we initially focus our attention on these models, given the substantial attention they have received in the literature regarding the estimation of sinusoidal parameters. The frequency parameter has been the subject of significant research attention due to its non-linear incorporation into a signal model. The entirety of research conducted in the field of data analysis has previously been based on a classical approach to statistical methodology. However, there has been a notable increase in interest in the Bayesian approach among scientists in a range of fields in recent times. In this analysis, unknown parameters are random variables distributed according to prior probabilities that express the analyst's degree of belief or ignorance. This allows for the derivation of statistical inferences about the parameters from their posterior distributions, which combine the analyst's beliefs with observations in a probabilistic framework. Nonetheless, the formulation of prior beliefs in the form of prior probability distributions on models under consideration, coupled with the calculation of the variables that underpin Bayesian model detection and parameter estimation, often presents a significant hurdle. These variables are typically integrals of substantial dimensions, lacking any closed-form analytical solution. These issues represent a significant challenge to the implementation of the Bayesian approach. Furthermore, the latest advances in computing have transformed Bayesian data analysis. Over the past few decades, we have seen the development of cutting-edge computational techniques that have led

to a new area of data analysis. In a few cases, for example, when the sinusoids are well separated, and many samples are available, suitable analytic approximations to these integrals can be performed. Moreover, these approximations are challenging to quantify and not valid in cases where the amount of available data is limited, and some sinusoids are close to each other. Some initial efforts to address this computational challenge using classical deterministic multiple integration techniques, as outlined in [1], [2], [3] and Monte Carlo methods, as presented in [3], [4], [5], [6], [7], [8], [9] have made. In addition, these approaches lack flexibility and become challenging to utilize when the dimension of the integrand becomes substantial.

The recent advances in the Bayesian literature, and in particular the advent of the reversible jump Markov chain Monte Carlo (RJMCMC) sampling [10], [11], [12] initially proposed by [11], have greatly simplified the determination of Bayesian models and the estimation of parameters. The method is based on the creation of a Markov chain, which can "jump" between models with parameter spaces of varying dimensions. In a Bayesian framework, the significant impact allows the calculation of probabilities for competing models. It can only explore the parameter space within the model space but can jump between plausible models. The RJMCMC sampler [11], [12], [13], [14] has trans-dimensional moves in addition to fixed-dimensional moves. Under certain conditions, these moves leave joint posterior probability density functions unchanged. Although RJMCMC has been extensively used in many applied model determination problems, its widespread applicability has been limited by the difficulty of achieving proposal moves between models that employ some notion of inter-model consistency, which facilitates good mixing across models. We therefore present a methodology that constructs moves between any models in model space in a general regression setting and illustrate its applicability.

In this paper, we examine the potential of combining the RJMCMC sampler with the SA algorithm [10], [15], [16]. We investigate the impact of varying proposal distributions on the effectiveness of this approach in detecting the number of sinusoids and estimating their parameters from noisy observations.

2 Mathematical Problem

Let $\mathbf{D}=\{d_1,d_2,...,d_N\}^t$ be a vector of N samples of an observed signal, which is a subset of a finite family of embedded models M_k, with $k\in\{0,...,k_{\max}\}$. The model M_k assumes that $\mathbf{D}$ can be written as a linear combination of k sinusoids corrupted by a zero mean white Gaussian noise $n_k(t)$ with a variance σ_k^2, as follows

$$M_k: d(t)=\begin{cases} n_k(t) & k=0 \\ \sum_{j=1}^{k} a_{c_j}^k \cos(\omega_j^k t)+a_{s_j}^k \sin(\omega_j^k t)+n_k(t) & k>0 \end{cases} \quad (1)$$

where $a_{c_j}^k$, $a_{s_j}^k$ and ω_j^k are the amplitudes and the radial frequencies of the j th sinusoids of the k th model signal, respectively. Without loss of generality, we assume that the frequencies $\omega_j^k\in(0,\pi)$ and $\omega_{j_1}^k\neq\omega_{j_2}^k$ for $j_1\neq j_2$. In vector-matrix form, we can rewrite the Eq. (1) for $k>0$ as

$$\mathbf{D}=\mathbf{G}(\boldsymbol{\omega}_k;\mathrm{t})\mathbf{a}_{2k}+\mathbf{n}_k, \quad (2)$$

where $\mathbf{a}_{2k}^T=(a_{c_1},a_{s_1},...,a_{c_k},a_{s_k})$ is the $2k$-dimensional vector of amplitudes, and $\mathbf{G}(\boldsymbol{\omega}_k;\mathbf{t})$ is a $(N\times 2k)$ matrix, defined by

$$\mathbf{G}(\boldsymbol{\omega}_k;\mathbf{t})=\begin{pmatrix} \cos(\omega_1 t_1) & \sin(\omega_1 t_1) & \cdots & \sin(\omega_k t_1) \\ \vdots & \vdots & \vdots & \vdots \\ \cos(\omega_1 t_N) & \sin(\omega_1 t_N) & \cdots & \sin(\omega_k t_N) \end{pmatrix}. \quad (3)$$

The subscript k that represents the number of sinusoids and the parameters $\boldsymbol{\theta}_k=\{\mathbf{a}_{2k},\boldsymbol{\omega}_k,\sigma_k^2\}$ in the model M_k are unknowns. Therefore, given the data set $\mathbf{D}$ our goal is to determine k and estimate $\boldsymbol{\theta}_k$ simultaneously. Although inference about two kinds of unknowns $(k,\boldsymbol{\theta}_k)$ is based on different logical principles Bayes paradigm offers an opportunity of a single logical framework briefly described in the following sections.

3 Bayesian Framework

Bayesian parameter inference task [10], [11], [15], [16], [17], [18], [19], [20], [21] is therefore to infer the probability over parameters $\boldsymbol{\theta}_k$ for the hypothesis or model M_k, given some data $\mathbf{D}$ from

experiment and capturing all relevant information I. This can be done within the setting of Bayes' Theorem which states

$$p\left(\boldsymbol{\theta}_k \middle| \mathbf{D}, M_k, I\right) = \frac{p\left(\mathbf{D}\middle|\boldsymbol{\theta}_k, M_k, I\right) p\left(\boldsymbol{\theta}_k \middle| M_k, I\right)}{p\left(\mathbf{D}\middle| M_k, I\right)}, \quad (4)$$

where $p\left(\boldsymbol{\theta}_k \middle| \mathbf{D}, M_k, I\right)$ is the posterior PDF, $p\left(\mathbf{D}\middle|\boldsymbol{\theta}_k, M_k, I\right)$ is the likelihood, $p\left(\boldsymbol{\theta}_k \middle| M_k, I\right)$ is the prior PDF and $p\left(\mathbf{D}\middle| M_k, I\right)$ is evidence. Bayes' theorem not only helps us to infer parameter distributions but, also provides a framework for model comparison. To compare models we take posterior odds of two models M_k and $M_{k'}$ by taking the ratio and cancelling the term $p\left(\mathbf{D}\middle| M_k, I\right)$ Thus

$$\frac{p\left(M_{k'}\middle|\mathbf{D}, I\right)}{p\left(M_k \middle| \mathbf{D}, I\right)} = \frac{p\left(\mathbf{D}\middle| M_{k'}, I\right) p\left(M_{k'}\middle| I\right)}{p\left(\mathbf{D}\middle| M_k, I\right) p\left(M_k \middle| I\right)} \quad (5)$$

Although inferences about two kinds of unknowns $(k, \boldsymbol{\theta}_k)$ are based on different logical principles Bayes paradigm offers an opportunity of a single logical framework briefly described in the following sections.

3.1 Prior Distributions and Likelihood

We set up a parameter space that is a countable union of subspaces of possible varying dimensionality:

$$\boldsymbol{\Theta} = \bigcup_{k=0}^{k_{\max}} \{k\} \times \Theta_k, \quad (6)$$

where $\Theta_0 = \Re^+$ corresponds to the case where the signal consists just of noise and $\Theta_k = \Re^{2k} \times \Omega_k \times \Re^+$ for $k \in \{0, 1, ..., k_{\max}\}$ with $\Omega_k = \{0, \pi\}^k$ and $k_{\max} = (N-1)/2$. In addition, we define that $\Omega = \bigcup_{k=0}^{k_{\max}} \{k\} \times \Omega_k$ where $\Omega_0 = \varnothing$. On this parameter space, we consider assignments of the prior PDF of all parameters by using the product rule of probability calculus:

$$p\left(k, \boldsymbol{\omega}_k, \mathbf{a}_k, \sigma_k^2\right) = p\left(k, \boldsymbol{\omega}_k, \mathbf{a}_k \middle| \sigma_k^2\right) p\left(\sigma_k^2\right)$$
$$= p(k)\, p(\boldsymbol{\omega}_k | k)\, p\left(\mathbf{a}_k \middle| \sigma_k^2\right) p\left(\sigma_k^2\right), \quad (7)$$

where σ_k^2 is a scale parameter, assumed to be distributed according to a conjugate inverse Gamma prior distribution, $\sigma_k^2 \approx \Gamma^{-1}\left(\nu_0/2, \gamma_0/2\right)$. When $\nu_0 = 0$ and $\gamma_0 = 0$ it turns out Jeffreys' uninformative prior [22], [23], [24] $p\left(\sigma_k^2\right) \propto 1/\sigma_k^2$ To facilitate the subsequent analysis, we propose the following prior distribution for $(k, \boldsymbol{\theta}_k)$:

$$p\left(k, \boldsymbol{\omega}_k, \mathbf{a}_k \middle| \sigma_k^2\right) = p(k)\, p(\boldsymbol{\omega}_k | k)\, p\left(\mathbf{a}_k \middle| \sigma_k^2\right)$$
$$= \frac{\Lambda^k \exp(-\Lambda)}{k!} \times \frac{I_{\Omega_k}(\boldsymbol{\omega}_k)}{\pi^k} \quad (8)$$
$$\times \left|2\pi\sigma_k^2 \boldsymbol{\Sigma}_k\right|^{-0.5} \exp\left(\frac{-\mathbf{a}_k \boldsymbol{\Sigma}_k^{-1} \mathbf{a}_k}{2\sigma_k^2}\right)$$

where $\boldsymbol{\Sigma}_k^{-1} = \delta^{-2} \mathbf{G}(\boldsymbol{\omega}_k, \mathbf{t})^T \mathbf{G}(\boldsymbol{\omega}_k, \mathbf{t})$ and $\delta \in \Re^+$ is an expected signal-to-noise ratio. If $k = 0$ it is assumed that $\mathbf{a}_0 \boldsymbol{\Sigma}_0^{-1} \mathbf{a}_0 = 0$ and $\left|2\pi\sigma_0^2 \Sigma_0\right|^{-0.5} = 1$. The first term in Eq. (8) is a prior PDF of k which is a truncated Poisson distribution with an expected number of sinusoids $\Lambda \in \Re^+$; the second term is a prior PDF of $\boldsymbol{\omega}_k$ which is also uniformly distributed on the condition that k is given and $\mathbf{I}_{\Omega_k}(k, \boldsymbol{\omega}_k)$ is an indicator function of the set Ω_k, defined in the following sense:

$$\mathbf{I}_{\Omega_k}(k, \boldsymbol{\omega}_k) = \begin{cases} 1 & \text{if } (k, \boldsymbol{\omega}_k) \in \Omega_k \\ 0 & \text{if } (k, \boldsymbol{\omega}_k) \notin \Omega_k \end{cases}. \quad (9)$$

The last term is a prior PDF of $\mathbf{a}_k$ which is a zero mean Gaussian with a covariance $\sigma_k^2 \boldsymbol{\Sigma}_k^{-1}$. Next, we consider the assignment of a functional form for the likelihood function, which is equivalent to the direct probability for data given the signal parameters. In the absence of a detailed knowledge of the noise distribution, other than that it has a finite variance σ_k^2, the application of the principle of maximum entropy [25] informs us that a Gaussian distribution is the most conservative choice for the direct probability of the data $\mathbf{D}$. Nevertheless, the application of the maximum entropy principle to assign a Gaussian distribution to the noise does not imply any assertion regarding its true sampling distribution. Instead, it merely represents a state of knowledge that is as uninformative as possible. Based on these considerations, the likelihood

function for our problem can be formulated as follows:

$$p(\mathbf{D}|k,\boldsymbol{\omega}_k,\mathbf{a}_k)=\left(2\pi\sigma_k^2\right)^{-\frac{N}{2}}\exp\left(-\frac{\chi^2}{2}\right), \quad (10)$$

where

$$\chi^2\left(\boldsymbol{\omega}_k,\mathbf{a}_k\right)=\frac{\left(\mathbf{D}-\mathbf{G}\left(\boldsymbol{\omega}_k,t\right)\mathbf{a}_k\right)^T\left(\mathbf{D}-\mathbf{G}\left(\boldsymbol{\omega}_k,t\right)\mathbf{a}_k\right)}{\sigma_k^2}. \quad (11)$$

3.2 Posterior Distribution

Bayesian statistical inference of k and $\boldsymbol{\theta}_k$ involves calculating their posterior PDF $p\left(k,\boldsymbol{\theta}_k|\mathbf{D}\right)$ expressed as:

$$\begin{aligned}p\left(k,\boldsymbol{\theta}_k|\mathbf{D}\right)&=p\left(k,\boldsymbol{\omega}_k,\mathbf{a}_k,\sigma_k^2|\mathbf{D}\right)\\&=p\left(\mathbf{D}|k,\boldsymbol{\omega}_k,\mathbf{a}_k,\sigma_k^2\right)\\&\quad\times p\left(a_k|\sigma_k^2,\boldsymbol{\omega}_k,k\right)\times p\left(\boldsymbol{\omega}_k|k\right)\\&\quad\times p\left(\sigma_k^2\right)\times p(k)\end{aligned}. \quad (12)$$

By using Equations (8) and (10) in (12) and simplifying the resulting expressions, we obtain the following posterior PDF for $k>0$:

$$\begin{aligned}p(k,\boldsymbol{\theta}_k|\mathbf{D})\propto&\left|2\pi\sigma_k^2\boldsymbol{\Sigma}_k\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{a}_k-\hat{\mathbf{a}}_k)^T\mathbf{M}_k^{-1}(\mathbf{a}_k-\hat{\mathbf{a}}_k)\right)\\&\times\exp\left(\frac{-1}{2\sigma_k^2}\left(\gamma_0+\mathbf{D}^T\mathbf{P}_k\mathbf{D}\right)\right)\\&\times\frac{\left(\frac{\Lambda}{\pi}\right)^k\mathbf{I}_{\Omega_k}(k,\boldsymbol{\omega}_k)}{k!}\times\left(\sigma_k^2\right)^{-\left(\frac{\nu_0}{2}+1\right)}\exp\left(-\frac{\gamma_0}{2\sigma_k^2}\right)\end{aligned}, \quad (13)$$

where

$$\begin{aligned}\mathbf{M}_k^{-1}&=\mathbf{G}^T(\boldsymbol{\omega}_k,t)\mathbf{G}(\boldsymbol{\omega}_k,t)+\boldsymbol{\Sigma}_k^{-1}\\\hat{\mathbf{a}}_k&=\mathbf{M}_k\mathbf{G}^T(\boldsymbol{\omega}_k,t)\mathbf{D}\end{aligned}, \quad (14)$$

and

$$\mathbf{P}_k=\mathbf{I}_N-\mathbf{G}^T\left(\boldsymbol{\omega}_k,t\right)\mathbf{M}_k\mathbf{G}\left(\boldsymbol{\omega}_k,t\right), \quad (15)$$

where $\mathbf{I}_N$ is a $\left(N\times N\right)$ dimensional identity matrix. By completing the square of Eq. (13), we integrate out with $\hat{\mathbf{a}}_k$:

$$p(\mathbf{a}_k\mid\mathbf{D},\boldsymbol{\omega}_k,\sigma_k^2)\propto\mathrm{N}(\hat{\mathbf{a}}_k,\sigma_k^2\mathbf{M}_k), \quad (16)$$

and then σ_k^2:

$$p(\sigma_k^2|\mathbf{D},\boldsymbol{\omega}_k)\propto\Gamma^{-1}(\frac{\nu_0+N}{2},\frac{\gamma_0+\mathbf{D}^T\mathbf{P}_k\mathbf{D}}{2}), \quad (17)$$

Finally, integration of Eq. (13) with respect to $\hat{\mathbf{a}}_k$ and σ_k^2 yields:

$$\begin{aligned}p\left(k,\boldsymbol{\omega}_k|\mathbf{D}\right)\propto&\left(\gamma_0+\mathbf{D}^T\mathbf{P}_k\mathbf{D}\right)^{-\frac{(N+\nu_0)}{2}}\\&\times\frac{\left(\frac{\boldsymbol{\Lambda}}{(1+\delta^2)\pi}\right)^k\mathbf{I}_{\Omega_k}(k,\boldsymbol{\omega}_k)}{k!}\end{aligned}. \quad (18)$$

Although considerable complexity reduction is made by integrating out the nuisance parameters $\left(\mathbf{a}_k,\sigma_k^2\right)$, Eq. (18) is still extremely complicated so that analytic methods cannot be used to calculate any statistics, and the model dimension is also not fixed. Therefore, the BI-RJMCMC -SA algorithm has been chosen to explore the posterior over the different spaces.

4 Bayesian Computation

To perform the Bayesian computation in extracting the parameters of interest from the posterior PDF given in Eq. (18), k is initially chosen at random from the uniform distribution $U\left(\left\{1,k_{\max}\right\}\right)$, where $k_{\max}=\left(N-1\right)/2$. Then, we introduced an algorithm, summarized in Table 1 (Appendix), which is based on two steps. The first step is called RJMCMC step which consists of a variety of Metropolis Hastings (MH) moves under different proposal distributions, [26], [27]. Once we get to update the hyper parameter δ^2 and the second step uses the SA algorithm to perform a global search in the joint space of the parameters, thereby surmounting the problem of local optima to a large extent.

4.1 RJMCMC Step

It allows the sampler to jump between spaces of different dimensionality. This is achieved by doing Markov chain sampling in spaces of varying dimensions.

RJMCMC unequivocally can change the dimension of the state space in a single move. This move is called a reversible jump [4], [11], [20], [28], [29] and allows the algorithm to switch back to the recent space with a later move. The reversibility of Markov chains is fundamental and

exploited in classical MH algorithms. A given distribution $\pi\left(k,\boldsymbol{\omega}_k\right)$ is easily ascertained as an invariant Markov distribution, [30]. Then our task is to make an inference about the joint distribution $\pi\left(k,\boldsymbol{\omega}_k\right)$ which is Bayesian in nature:

$$\pi\left(k,\boldsymbol{\omega}_k\right)=p(k,\boldsymbol{\omega}_k\left|\mathbf{D}\right) . \quad (19)$$

The constant of proportionality is not required, as it cancels out of the numerator and denominator of the acceptance ratio. As [8] notes, the Bayesian formulation, when used in conjunction with RJMCMC, provides a logical approach to making combined inferences about $\left(k,\boldsymbol{\omega}_k\right)$.

Here, we implement the Markov transition from a current state ($\mathbf{X}$) by first proposing a new state ($\mathbf{X}'$), drawn from proposal distribution $q\left(\mathbf{X}'\left|\mathbf{X}\right.\right)$. We accept the move with a probability $\alpha\left(\mathbf{X},\mathbf{X}'\right)$, and if rejected, the chain remains at the current state. We compute the acceptance probability,

$$\alpha\left(\mathbf{X},\mathbf{X}'\right)=\min\left\{1,A\left(\mathbf{X},\mathbf{X}'\right)\right\}, \quad (20)$$

which includes the Jacobian term to satisfy the detailed balance [6], [8], [11], [23], [26] within each jump:

$$A\left(\mathbf{X},\mathbf{X}'\right)=\frac{\pi\left(\mathbf{X}'\right)q(\mathbf{X}\left|\mathbf{X}'\right)}{\pi\left(\mathbf{X}\right)q\left(\mathbf{X}'\left|\mathbf{X}\right.\right)}\left|\frac{\partial\left(\mathbf{X}'\right)}{\partial\left(\mathbf{X}\right)}\right|, \quad (21)$$

where $\mathbf{X}=\left(k,\boldsymbol{\omega}_k\right)$ and $\mathbf{X}'=\left(k',\boldsymbol{\omega}_{k'}^{'}\right)$. Although convenient choices of mappings may be far from optimal and identifying a potential mapping may be difficult in most of applications, various approaches have already been proposed to both automate and improve efficiency between model proposals [15], [16], [17], [18], [23].

The exploration of different moves, either deterministically or randomly, is now possible within a range of $\left(k,\boldsymbol{\omega}_k\right)$. Consequently, a candidate move is selected at random, resulting in a transition density for the simulated Markov chain that is a mixture of different transition densities associated with various move types (birth, death, split, merge, and update), with equal probabilities. A random walk is then performed.

4.1.1 Birth-Death Move

Let us suppose that the current state of the Markov chain is in $\left\{k\right\}\times\Omega_k$. We now propose a new frequency ω' chosen at random on the interval $\left(0,\pi\right)$, set up $\boldsymbol{\omega}_{k+1}=\left\{\boldsymbol{\omega}_k,\omega'\right\}$ and then evaluate the acceptance ratio $\alpha_{k,k+1}$:

$$\alpha_{k,k+1}=\min\left\{1,r_{birth}\right\}, \quad (22)$$

where

$$r_{birth}=\left(\frac{\gamma_0+\mathbf{D}^T\mathbf{P}_k\mathbf{D}}{\gamma_0+\mathbf{D}^T\mathbf{P}_{k+1}\mathbf{D}}\right)^{\frac{(\nu_0+N)}{2}}\frac{1}{\left(k+1\right)\left(1+\delta^2\right)} . \quad (23)$$

If a random number u generated from a uniform distribution $U\left(\left\{0,1\right\}\right)$ satisfies $u\le\alpha_{k,k+1}$ then the new state of the Markov chain becomes $\left(k+1,\boldsymbol{\omega}_{k+1}\right)$; otherwise, we stay at the current state $\left(k,\boldsymbol{\omega}_k\right)$.

Let us suppose that the current state of the Markov chain is in $\left\{k+1\right\}\times\Omega_{k+1}$. We then choose a sinusoid with a label $l\in\left\{1,2,...,k+1\right\}$ at random among the $(k+1)$ existing sinusoids and remove it. Thus, we get a new state of the Markov chain $\left(k,\boldsymbol{\omega}_k\right)$ and accept it if a random number u generated from a uniform distribution $u\approx U\left(\left\{0,1\right\}\right)$ satisfies $u\le\alpha_{k+1,k}=\min\left\{1,r_{birth}^{-1}\right\}$; otherwise, we stay at the current state $\left(k+1,\boldsymbol{\omega}_{k+1}\right)$.

4.1.2 Split-Merge Move

These moves are motivated when two sinusoids are closely spaced in frequency. In split move, we choose a sinusoid with a label $l\approx U\left(\left\{1,k\right\}\right)$ among the existing ones. Then, the l th sinusoid with parameters $\left\{a_{c_l},a_{s_l},\omega_l\right\}$ is split into two sinusoids with parameters:

$$\begin{aligned}a_{c_1}^{*}&=\varepsilon_1\sqrt{u_2u_1(a_{c_l}^2+a_{s_l}^2)}\\ a_{s_1}^{*}&=\varepsilon_2\sqrt{(1-u_2)u_1(a_{c_l}^2+a_{s_l}^2)}\\ \omega_1^{*}&=\omega_l-u_4\sigma_{\omega_l}\sqrt{\frac{(a_{c_2}^{*2}+a_{s_2}^{*2})}{(a_{c_1}^{*2}+a_{s_1}^{*2})}}\end{aligned}, \quad (24)$$

and

$$a_{c_2}^* = \varepsilon_1 \sqrt{u_3(1-u_1)(a_{c_l}^2 + a_{s_l}^2)}$$
$$a_{s_2}^* = \varepsilon_2 \sqrt{(1-u_3)(1-u_1)(a_{c_l}^2 + a_{s_l}^2)} \quad (25)$$
$$\omega_2^* = \omega_l + u_4 \sigma_{\omega_l} \sqrt{\frac{(a_{c_1}^{*2} + a_{s_1}^{*2})}{(a_{c_2}^{*2} + a_{s_2}^{*2})}}$$

where σ_{w_l} and $\varepsilon_l \left(l = 1,2\right)$ are a predetermined constant. Once $\boldsymbol{\omega}_1^*$ and $\boldsymbol{\omega}_2^*$ have been sampled. If there is no other frequency located between them then the move is rejected; otherwise, it is accepted if

$$u \approx U\left(\{0,1\}\right) < A_{k,k+1} = \min\left\{1, r_{split}\right\} \quad (26)$$

with

$$r_{split} = (1+\delta^2)^{-\frac{1}{2}} \left(\frac{\mathbf{D}^T \mathbf{P}_k \mathbf{D}}{\mathbf{D}^T P_{k+1} \mathbf{D}} \right)^{N/2} \varsigma^*, \quad (27)$$

where $\zeta^* = \left| \partial\left(\omega_1^*, \omega_2^*\right) / \partial\left(\omega_{l,}, u\right) \right|$.

We choose a label $l \approx U\left(\{1, k_{\max}\}\right)$ and $u \approx U\left(\{0,1\}\right)$. Then we get two sinusoids with parameters $\left\{a_{c_{l-1}}, a_{s_{l-1}}, \omega_{l-1}\right\}$ and $\left\{a_{c_l}, a_{s_l}, \omega_l\right\}$ so that we merge them into one sinusoid with parameters $\left\{a_{c_l}^*, a_{s_l}^*, \omega_l^*\right\}$:

$$a_{c_l}^* = \varepsilon_1 \sqrt{u_0(a_{c_{l-1}}^2 + a_{s_{l-1}}^2 + a_{c_l}^2 + a_{s_l}^2)},$$
$$a_{s_l}^* = \varepsilon_2 \sqrt{(1-u_0)(a_{c_{l-1}}^2 + a_{s_{l-1}}^2 + a_{c_l}^2 + a_{s_l}^2)} \quad (28)$$

and

$$\omega_l^* = \frac{(a_{c_{l-1}}^2 + a_{s_{l-1}}^2)\omega_{l-1} + (a_{c_l}^2 + a_{s_l}^2)\omega_l}{(a_{c_{l-1}}^2 + a_{s_{l-1}}^2) + (a_{c_l}^2 + a_{s_l}^2)}. \quad (29)$$

One can easily check that (25) and (26) are compatible with (28) and (29) so that it is accepted if:

$$u \approx U\left(\{0,1\}\right) < A_{k+1,k} = \min\left\{1, r_{split}^{-1}\right\}. \quad (30)$$

4.1.3 Update Move

This move does not entail modifying the dimensions of the model; rather, it necessitates an iteration of hybrid MCMC samplers. In the context of our application, the target distribution represents a full conditional distribution of a frequency:

$$p(\omega'_{j,k} \left| \mathbf{D}, \omega_{j,k} \right.) \propto \left(\gamma_0 + \mathbf{D}^T \mathbf{P}_k \mathbf{D}\right)^{-\frac{(N+v_0)}{2}} \mathbf{I}_{\Omega}(\boldsymbol{\omega}_k) \quad (31)$$

As with all MCMC methods that use the Metropolis-Hastings algorithm, suitable proposal distributions $q(\omega'_{j,k} \left| \omega_{j,k} \right.)$ for the proposed transitions must be chosen. The efficiency of the reversible jump sampler is highly dependent on this choice.

A cautious proposal distribution generating small steps will typically be well received but will nevertheless progress slowly. Conversely, a bold proposal distribution generating large steps will often result in moves from the body to the tails of the distribution, leading to a low acceptance ratio and a low probability of acceptance. However, determining an appropriate width for the proposal distribution of each frequency is a time-consuming process.

To circumvent this, we have opted to perform a Metropolis-Hasting step [6], [7], [8], [9] and [10] with four different proposal distributions, as outlined below, with an equal probability.

The first proposal distribution of $\omega'_{j,k}$ which is independent from the current state $\omega_{j,k}$, is defined in the form:

$$q_1\left(\omega'_{j,k} \left| \omega_{j,k} \right.\right) \propto \sum_{l=0}^{N_p-1} p_l \mathbf{I}_{(l\pi/N_p,(l+1)\pi/N_p)}(\omega'_{j,k}), \quad (32)$$
$$(j = 1,...,k)$$

where p_l is the value of the squared modulus of the Fourier transform of the observations $\mathbf{D}$ at frequency $(l\pi / N_p)$. We will take $N_p = N$ but, $N_p > N$ can also be used to improve the interpolation of the Fourier Transform via zero padding. Choosing a new frequency $\omega'_{j,k}$ independent from $\omega_{j,k}$ in the regions where the modulus of the Fourier transform has high values helps us to build the regions of interest of the posterior distribution quickly.

The second proposal distribution is, in common, defined as a normal distribution with the mean which is the current estimated values $\omega_{j,k}$ and the variance which is a Cramer-Rao lower bond [18], [19] defined by:

$$\sigma_{\omega_{j,k}}^2 \cong \frac{\sigma_k^2}{(a_{2j-1}^2 + a_{2j}^2) N(N^2 - 1)}, \quad (33)$$

that is a lower limit to the variance for a measurement of the current frequency $\omega_{j,k}$:

$$q_2(\omega'_{j,k} \left| \omega_{j,k} \right.) \propto \mathrm{N}(\omega_{j,k}, \sigma^2{}_{\omega_{j,k}}). \quad (34)$$

It can be demonstrated that the square root of the value in question generates a natural scale size of the search space around the estimated value of $\sigma^2_{\omega_{j,k}}$. Given that the search space is relatively smooth, it can be expected that superior solutions will be found near those that are already optimal $\omega_{j,k}$. This random walk with a normally distributed step size leads to the exploration of the posterior distribution on a local scale and ensures the irreducibility of the Markov chain.

If all estimated frequencies are sorted into ascending order $0<\omega_{1,k}<\omega_{2,k}<...<\omega_{k,k}<\pi$, one then assigns a broad uninformative Gaussian prior to each frequency by specifying its low and high value. From this range, a mean of any frequency $\omega_{j,k}$ is computed by $\mu_{\omega_{j,k}}=(\omega^{high}_{j,k}+\omega^{low}_{j,k})/2$ and the standard deviation of the Gaussian prior is set so the entire interval, low to high, represents a three-standard deviation of the interval $\sigma_{\omega_{j,k}}=(\omega^{high}_{j,k}-\omega^{low}_{j,k})/3$. Then the third proposal distribution can be defined by an uninformative Gaussian prior probability for the frequency $\omega_{j,k}$ in the following form:

$$q_3(\omega'_{j,k}\left|\omega_{j,k}\right.)\propto \mathrm{N}(\mu_{\omega_{j,k}},\sigma^2_{\omega_{j,k}}), \qquad (35)$$

where $\omega'_{j,k}\in(\omega^{low}_{j,k},\omega^{high}_{j,k})$. By making a discretization of the Langevin diffusion [30] using a first-order Euler approximation, we get:

$$\begin{aligned}&\mu=\omega_{j,k}+\sigma^2_{\omega_{j,k}}\nabla\log\left(p\left(k,\omega_{j,k}\right)\right);\\&q_4\left(\omega^*_{j,k}\left|\omega_{j,k}\right.\right)\propto \mathrm{N}\left(\mu,\sigma_{\omega_{j,k}}\right)\end{aligned}, \qquad (36)$$

where $\omega^*_{j,k}\in(\omega^{low}_{j,k},\omega^{high}_{j,k})$. According to (5), the target distribution becomes a full conditional PDF of $\omega_{j,k}$. Then we get:

$$\alpha_{k,k'}=\min\left\{1,\left(\frac{\mathbf{D}^T\mathbf{P}_k\mathbf{D}}{\mathbf{D}^T\mathbf{P}^*_{k'}\mathbf{D}}\right)^{N/2}\frac{q_l\left(\omega_{j,k}\left|\omega^*_{j,k'}\right.\right)}{q_l\left(\omega^*_{j,k'}\left|\omega_{j,k}\right.\right)}\right\}, \qquad (37)$$

$(l=1,...,4)$

where

$$\mathbf{P}^*_k=\mathbf{P}_k\Big|_{\boldsymbol{\omega}_k=\boldsymbol{\omega}^*_k},\ \mathbf{M}^*_k=\mathbf{M}_k\Big|_{\boldsymbol{\omega}_k=\boldsymbol{\omega}^*_k} \qquad (38)$$

and

$$\boldsymbol{\Sigma}^*_k=\boldsymbol{\Sigma}_k\Big|_{\boldsymbol{\omega}_k=\boldsymbol{\omega}^*_k}. \qquad (39)$$

4.2 Simulated Annealing

Our hierarchical Bayesian formulation demonstrates that traditional model selection criteria [4], [6], [26] within a penalized likelihood framework correspond to particular hyper-parameter choices. In other words, the prior choices can be calibrated in such a way that the issue of model selection within the context of penalized likelihood is transformed into a problem of model selection based on posterior probabilities. Once the calibration issue has been resolved, we proceed with maximum likelihood estimation using the aforementioned model selection criteria. This allows us to maximize the calibrated posterior distribution. To achieve this objective, we propose a simulated annealing algorithm that employs the homogeneous reversible jump MCMC kernel [19], [20], [22], [28], [31] as a proposal. This approach has the advantage of allowing an arbitrary model order to be used as a starting point, with the algorithm performing dimension jumps until the "true" model order is identified. This eliminates the need for a more expensive task, such as running a fixed-dimension algorithm for each possible model order and subsequently selecting the best model. In Mathematical terms, this estimate is given by

$$M_s=\underset{M_k:k\in\{0,k_{\max}\}}{\arg\min}\left\{-\log p\left(\mathbf{D}\left|k,\hat{\boldsymbol{\theta}}\right.\right)+P\right\}, \qquad (40)$$

where $\hat{\boldsymbol{\theta}}_k=\left\{\hat{\mathbf{a}}_k,\hat{\boldsymbol{\omega}}_k,\sigma^2_k\right\}$ is esimates of $\boldsymbol{\theta}_k$ for the model M_k. P is the penalty term that depends on the model order. In the case of Gaussian noise, $P_{AIC}=\xi$ and $P_{BIC}=P_{MDL}=\frac{\xi}{2}\log(N)$, where ξ denotes the number of model parameters. The above criteria are driven by a few factors, including the fact that the Akaike information criteria (AIC) [4] is founded upon anticipated information, the Bayesian information criteria (BIC) is an asymptotic Bayes factor, and the Minimum description length (MDL) necessitates the assessment of the minimal data transmission requirements [6], [10], [22].

Consequently, the penalized likelihood framework can be interpreted as a problem of maximizing the joint posterior distribution $p\left(k,\boldsymbol{\omega}_k\left|\mathbf{D}\right.\right)$. Effectively, this maximum posterior (MAP) estimate can be obtained as follows:

$$\left(k,\boldsymbol{\omega}_k\right)_{MAP} = \underset{k,\boldsymbol{\omega}_k\in\Omega_k}{\arg\max}\; p\left(k,\boldsymbol{\omega}_k\left|\mathbf{D}\right.\right) = \underset{k,\boldsymbol{\omega}_k\in\Omega_k}{\arg\max}\left\{\left(\mathbf{D}^T\mathbf{P}_k^*\mathbf{D}\right)^{-\frac{N}{2}}\exp\left(-P\right)\right\}. \quad (41)$$

The simulated annealing method involves simulating a non-homogenous Markov chain whose invariant distribution at iteration i is no longer equal to $\pi\left(k,\boldsymbol{\omega}_k\right)$, but to:

$$\pi\left(k,\boldsymbol{\omega}_k\right)\propto\pi^{\frac{1}{T_i}}\left(k,\boldsymbol{\omega}_k\right), \quad (42)$$

where T_i is a decreasing cooling schedule with $\lim_{i\to\infty}T_i=0$. As with MH method, the simulated annealing method with distribution $\pi\left(\mathbf{X}\right)$ and proposal distribution $q\left(\mathbf{X}'\left|\mathbf{X}\right.\right)$ involves sampling a candidate value $\mathbf{X}'$ given the current value X. The Markov chain moves towards X' with a probability

$$A_{SA}=\min\left\{1,\frac{\pi^{\frac{1}{T_i}}\left(\mathbf{X}'\right)q(\mathbf{X}\left|\mathbf{X}'\right.)}{\pi^{\frac{1}{T_i}}\left(\mathbf{X}\right)q(\mathbf{X}'\left|\mathbf{X}\right.)}\right\}, \quad (43)$$

otherwise, it remains equal to $\mathbf{X}$. It is of critical importance to select an effective proposal distribution to ensure the development of an efficient algorithm. A homogeneous transition kernel $K\left(\mathbf{X},\mathbf{X}'\right)$ that satisfies the following reversibility property is a key component in this process [21], [32], [33]:

$$\pi\left(\mathbf{X}'\right)K\left(\mathbf{X}',\mathbf{X}\right)=\pi\left(\mathbf{X}\right)K\left(\mathbf{X},\mathbf{X}'\right) \quad (44)$$

It follows that:

$$A_{SA}=\min\left\{1,\frac{\pi^{1/T_i-1}\left(\mathbf{X}'\right)}{\pi^{1/T_i-1}\left(\mathbf{X}\right)}\right\}. \quad (45)$$

These candidates are randomly accepted according to an acceptance ratio which ensures reversibility and thus invariance of the Markov chain with respect to the posterior distribution. Here, the chain must move across subspaces of different dimensions, and therefore the proposal distributions are more complex. In accordance with a set of proposal distributions, a random selection of candidates is made and accepted in accordance with an acceptable ratio. This provides reversibility and invariance of the Markov chain related to the posterior PDF in Eq. (5). For further details, please refer to [10], [11]. For the sake of simplicity, we will drop the superscript from all variables and assume that a candidate move is selected at random. The resulting transition distribution of the simulated Markov chain will then be a mixture of the different transition distributions related to the different move types, with equal probabilities.

5 Computer Simulations

This section will examine how effectively the previously discussed algorithm can summarise variable-dimensional posterior probability density functions (PDFs). These are the types of PDFs that may be encountered when undertaking joint detection and estimation of sinusoids in white Gaussian noise [34], [35], [36]. The only additional information required by the algorithm is a specification of a few other parameters, such as ν_0,γ_0,Λ and δ^2, which we will refer to as hyperparameters. The values of υ_0 and γ_0 can be fixed using a priori information about σ_k^2. As there is no specific a priori information about σ_k^2, we set $\nu_0=0$ and $\gamma_0=0$ so that $p(\sigma_k^2)\propto\frac{1}{\sigma_k^2}$ is known as an uninformative Jeffreys' prior [24]. As discussed in [8], [25], it has been demonstrated that by calibrating the priors in the hierarchical Bayesian formulation, by treating δ^2 and Λ as fixed quantities instead of as random variables. We may begin by assuming that the hyperparameters δ^2 is fixed to a particular value, say

$$\bar{\delta}^2=\mu_{IG\left(\alpha_{\delta^2},\beta_{\delta^2}\right)} \quad (46)$$

and

$$\Lambda=\left(1+\bar{\delta}^2\right)\pi\exp\left(-\left(1+\bar{\delta}^2\right)\right), \quad (47)$$

where $\alpha_{\delta^2}=2$ and $\beta_{\delta^2}=max\left\{N,k^2\right\}$. Simulations have demonstrated that the selection of Λ has no impact on the outcome of model order selection and parameter estimation. However, it does influence the duration of the transient state. For further details, please refer to the following papers [5], [9], [10].

The transition kernel of the simulated Markov chain is a mixture of the associated transition kernels. This is determined by selecting one of the candidate moves at each iteration, which can be birth, death, merge, split or update. The probabilities for choosing these moves are b_k, d_k, s_k, m_k and

u_k respectively, such that $b_k + d_k + s_k + m_k = 1$ with respect to all $k \in (0, k_{\max})$. A move is performed if the algorithm accepts it. For $k = 0$ the death, split and merge moves are impossible, so that $d_0 = s_0 = m_0 = 0$. The merge move is also not permitted for $k = 1$, that is $m_1 = 0$. For $k = k_{\max}$, the birth and split moves are not allowed and therefore $b_{k_{\max}} = s_{k_{\max}} = 0$.

The algorithm summarized in Table 1 (Appendix) needs some initial starting configurations k, $\boldsymbol{\omega}_k$ and T. An initial value for k is chosen at random from the interval $\{1, k_{\max}\}$, where $k_{\max}$ is chosen as $k_{\max} = 10$. To start iterations, the initial values for $\boldsymbol{\omega}_k$ are selected from k largest values of Fourier Power spectral densities [20], [34] because they are more likely to occur. As is well documented, SA constitutes a general-purpose global optimization algorithm that makes use of Monte Carlo calculations in a multidimensional solution space formed by the combination of all permissible values of the free model parameters. Then it will explore that space in a series of n_s steps, starting from a given point, $\{k, \boldsymbol{\omega}_k\}$ which represents a *state* of the values of the model parameters at any iteration. In each step of the SA, a *transition* to a new state $\{k^*, \boldsymbol{\omega}_k^*\}$ is proposed by *RJSA* step and the algorithm decides whether to accept the transition following a probabilistic criterion given in (45). On the other hand, the temperature T_i called control parameter varies from high to low values towards the end. The whole cycle is then repeated $n_T (\cong 100)$ times, after which T_i is gradually decreased by annealing (or cooling) schedule which is the heart of the SA algorithm. Although there are various cooling schedules suggested by different researchers [4], [7] for lowering, T_i we choose:

$$T_{i+1} = \eta_T T_i \exp\{-(\eta_T - 1)\kappa^{1/k}\}, \quad (48)$$

where κ represents several iterations and $\eta_T \in (0,1)$ is a constant. Finally, the algorithm terminates when average function value of the sequences of the points after each $n_s \times n_T$ step cycle reaches a stable state:

$$\left| p_i\left(\mathbf{D}\middle|k, \boldsymbol{\omega}_k\right) - p_{i-r}\left(\mathbf{D}\middle|k^*, \boldsymbol{\omega}_k^*\right)\right| \le \varepsilon$$
$$\left| p_i\left(\mathbf{D}\middle|k, \boldsymbol{\omega}_k\right) - p_{opt}\left(\mathbf{D}\middle|k^*, \boldsymbol{\omega}_k^*\right)\right| \le \varepsilon \quad , \quad (49)$$
$$(r = 1, ..., 4)$$

where ε is defined by user, $p_{opt}\left(\mathbf{D}\middle|k^*, \boldsymbol{\omega}_k^*\right)$ the last current optimum value and r indicates the last four successive iterations.

The proposed algorithm outlined in Section 4, was implemented in Mathematica [37], a software platform offering a comprehensive range of programming techniques and a vast array of built-in functions. This resulted in a significantly more concise computer code than those written in comparable languages, such as C or FORTRAN. The algorithm's performance was evaluated on a workstation equipped with four processors.

In this section, we present experimental results which demonstrate the performance of the selection rules. For the purposes of this investigation, three experiments were considered. The first experiment involved the generation of data which was in accordance with:

$$d(t) = a_1 \cos(\omega_1 t + \phi_1) + a_2 \cos(\omega_2 t + \phi_2) + e(t), \quad t \in Z^N \quad , (50)$$

where

$$a_1 = a_2 = \sqrt{20}, \ \phi_1 = 0 \text{ rad.}, \phi_2 = \frac{\pi}{4} \text{rad.}$$
$$\omega_1 = 2\pi 0.2, \ \omega_1 = 2\pi\left(0.2 + \frac{1}{N}\right) \quad , (51)$$

and $N = 64$. Throughout the experiment, the SNR [38] defined by:

$$SNR = 10 \log_{10} \frac{a_i^2}{2\sigma^2}, \quad (52)$$

was varied from -3 to 10 dB in steps of 1 dB. The noise sequences were generated according to a Gaussian density function with σ^2 appropriately chosen to yield the required SNR.

We let the reversible jump SA sampler to run and go through 10 thousand iterations. The computer output of simulations, illustrated in Table 2 (Appendix). The estimated parameter values are quoted as *estimated value* ± *standard deviation* and show that all parameters are clearly recovered within the calculated accuracy. Figure 1(a) (Appendix) shows predicted signal and data samples

with errors that indicate uncertainty in the data points. In Figure 1 (b) (Appendix), it is seen that the proposed algorithm clearly separates the noise in data. In addition, Figure 1(c) (Appendix) shows a comparison of Fourier and Bayes spectral density [15], [39], [40], [41]. A comparison of these two spectral densities indicates that frequencies obtained by the proposed algorithm have separated the sinusoids very well although the separation of the sinusoids is less than the Rayleigh resolution [4], [42]. However, as the number of samples was increased its performance was improved. In the second experiment the data $\mathbf{D}$, were represented three closely spaced sinusoids and generated by

$$d(t) = a_1 \cos(\omega_1 t + \phi_1) + a_2 \cos(\omega_2 t + \phi_2) + a_3 \cos(\omega_3 t + \phi_3) + e(t), \quad t \in Z^N \quad (53)$$

where

$$a_1 = a_3 = \sqrt{20}, \; a_2 = \sqrt{6.3246,}$$
$$\phi_1 = 0 \text{ rad}, \; \phi_2 = \frac{\pi}{4}\text{rad}, \; \phi_3 = \frac{\pi}{3}\text{rad}, \quad (54)$$
$$\omega_1 = 2\pi(0.2), \; \omega_2 = 2\pi\left(\omega_1 + \frac{1}{N}\right),$$

and $\omega_3 = 2\pi\left(\omega_1 + \frac{2}{N}\right)$. In a similar way, we produced data samples $N = 64$ shown in Figure 2(a) (Appendix) and ran the Mathematica code of this proposed algorithm again for this experiment. After 10 thousand iterations, Figure 2(b) (Appendix) shows the histogram and the probability density of approximated noise and its exact distribution. The best estimates of parameters for this frequency signal model are tabulated in Table 3 (Appendix). Once again, all the frequencies have been well resolved. Even the second frequencies which are too close not to be separated by the Fourier power spectral density shown in Figure 2(c) (Appendix). With the Fourier spectral density when the separation of two frequencies is less than the Nyquist step, defined as $2\pi / N$, the two frequencies are indistinguishable. This is simply because there are no sample points in between the two frequencies in the frequency domain. If $|\omega_1 - \omega_2| > 2\pi / N$ theoretically, the two frequencies can then be distinguished. If $|\omega_1 - \omega_2|$ is not large enough, the resolution will be very poor. Therefore, it is hard to tell where the two frequencies are located. This is just the inherent problem of the discrete Fourier power spectral density. In this example two frequencies are separated by 0.01, which is less than the Nyquist step size so that there is no way by using Fourier power spectral density that one can resolve these closed frequencies. However, Bayesian power spectral density shown in Figure 2(c) (Appendix) gives us very good results with high accuracy. These results are also consistent with those obtained by. [4], [6], [7] and demonstrate the advantages of BI-RJMCMC-SA. To evaluate the efficacy of the algorithm, it was deployed in a scenario utilizing a synthetic data set derived from a multiple harmonic frequency model signal, which was also employed in reference, [40]:

$$d(t) = \cos(0.1t+1) + 2\cos(0.15t+2) + 5\cos(0.3t+3) + 2\cos(0.31t+4) + 3\cos(t+5) + e(t) \quad (55)$$

The symmetric time interval $\{-256, 256\}$ is traversed in discrete steps. Figure 3(a) (Appendix) depicts noisy data. The results presented in Table 4 (Appendix) are expressed as estimated *values* ± *standard deviations.* As anticipated, all parameters are accurately estimated within the specified accuracy limits. A comparison of the current results with those obtained by [19] and [40] reveals that all frequencies are well resolved, including the third and fourth, which are in relatively closed proximity and therefore require separation. The typical representation of the results from a spectral analysis is in the form of power spectral density. Figures 3(b) (Appendix) shows the Fourier and Bayesian power spectral densities for this model signal, respectively. The Fourier power spectral density exhibits a mere four peaks, whereas the Bayesian power spectral density displays five peaks, which aligns with the number of frequencies present in the model signal.

The algorithm proposed here automatically estimates the number of sinusoids in data. Computer simulations demonstrate that there is no limit to the number of sinusoids. If a scheme is devised that allows the reachability of any model from any point in the state space, the RJMCMC-SA sampler will move through it and spend most of its time with a model or set of models that are the most likely to occur.

6 Conclusion

We have presented a straightforward and direct approach for the simultaneous detection of the number of sinusoids and the estimation of their parameters from noisy data. Our method employs a Bayesian inference in conjunction with RJMCMC-SA sampling, thereby providing a robust solution to the challenging problem of estimating the number of

sinusoids and their parameters from complex, noisy data sets. A concise overview of the underlying theory and illustrative examples of its practical applications are also provided. The results indicate that, despite the computational intensity, Bayesian RJMCMC-SA sampling is a powerful methodology for signal processing. Nevertheless, there remains a certain reluctance to employ MCMC algorithms, largely due to concerns pertaining to programming and reliability-convergence. It is thus recommended that further investigations be conducted to address this reluctance and facilitate further advances in the Bayesian exploration of complexity. Additionally, these analyses are particularly well suited for computation in the Mathematica program.

References:

[1] P.M. Djurić, A Model Selection Rule for Sinusoids in White Gaussian Noise, *IEEE Trans. Signal Processing,* Vol. 44, 7, 1996, pp. 1744–1751. DOI**:** 10.1109/78.510621.

[2] P.M. Djurić, Model Selection Based on Bayesian Predictive Densities and Multiple Data Records, *IEEE Transaction on Signal Processing,* Vol. 42, 7, 1994, pp. 1685-1699. DOI:10.1109/78.298276.

[3] W.K. Hastings, Monte Carlo Sampling Methods using Markov Chains and Their Applications, *Biometrika*, 1970, Volume 57, Issue 1, pp. 97-109. DOI: 10.2307/2334940.

[4] C. Andrieu, A. Doucet, Joint Bayesian Model Selection and Estimation of Noisy Sinusoids via Reversible Jump MCMC, *IEEE Transaction on Signal Processing,* Vol 47, 10, 1999, pp. 2667-2676. DOI**:** 10.1109/78.790649.

[5] C. Andrieu, Y.F. Atchade, On the Efficiency of Adaptive MCMC Algorithms, *Electronic Communications in Probability*, Vol 12, 2007, pp. 336-349. DOI: 10.1214/ECP.v12-1320.

[6] C. Andrieu, P.M. Djurić, A. Doucet, Model Selection by MCMC Computation, *Signal Processing*, Vol 81, 1, 2001, pp. 19-37. DOI: 10.1016/S0165-1684(00)00188-2.

[7] C. Andrieu, E. Barat, A. Doucet, Bayesian Deconvolution of Noisy Filtered Point Processes, *IEEE Transaction on Signal Processing,* Vol. 49, 2001, pp. 134-146. DOI: 10.1109/78.890355.

[8] C. Andrieu, N. D. Freıtase. M. I. Jordan and A. Doucet, An Introduction to MCMC for Machine Learning, Machine Learning, Vol. 50, 1,2003, pp. 5–43. DOI: 10.1023/A:1020281327116

[9] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, Equations of State Calculations by Fast Computing Machine, *Journal of Chemical Physics,* Vol. 21, 1953, pp. 1087-1091. DOI: 10.1063/1.1699114.

[10] Y. Fan., G.W. Peters, S.A. Sisson, Automating and Evaluating Reversible Jump MCMC Proposal Distributions, *Statistical Computing,* Vol 19, 4, 2009, pp. 469-421. DOI: 10.1007/s11222-008-9101-z.

[11] P.J. Green, Reversible Jump MCMC Computation and Bayesian Model Determination, *Biometrika,* Vol. 82, 4, 1995, pp. 711–732. DOI: 10.1093/biomet/82.4.711.

[12] P.C. Gregory, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with `Mathematica', Cambridge University Press, Cambridge, UK, 2005. DOI: 10.1017/CBO9780511791277.

[13] A.E. Gelfand, A.F.M. Smith, Sampling Based Approaches to Calculating Marginal Densities, *Journal of American Statistical Association*, Vol. 85, 1995, pp. 398-409. https://doi.org/10.2307/2289776.

[14] A. Jasra, C.C. Holmes, D.A. Stephens, Markov Chain Monte Carlo Methods and Label Switching Problem in Bayesian Mixture Modelling, *Statistical Science*, Vol. 20, 2005, pp. 50-67. DOI: 10.1214/088342305000000016.

[15] F. Al-Awadhi, M.A. Hurn, C. Jennison, Improving the Acceptance Rate of Reversible Jump MCMC Proposals, *Statistics and Probability Letters***,** Vol. 69, 2, 2004, pp. 189-198. DOI: 10.1016/j.spl.2004.06.025.

[16] D. Üstündag, M. Cevri, Bayesian Parameter Estimation of Sinusoids with Simulated Annealing, *8th Wseas International Conference on Signal Processing, Computational Geometry and Artificial Vision (ISCGAV'08)*, Rhodes, Greece, 2008, pp.106-112. ISBN**:** 978-960-6766-95-4.

[17] D. Üstündag, M. Cevri., Estimating of Sinusoids from Noisy Data using Bayesian Inference with Simulated Annealing, *Wseas Transactions on Signal Processing*, Vol. 4**,** 8, 2008, pp. 432-441. DOI: 10.5555/1481986.1481988.

[18] K. Copsey, N. Gordon, A. Marrs, Bayesian Analysis of Generalized Frequency-Modulated Signals, *IEEE Transactions on Signal Processing*, Vol. 50, 3, 2002, pp. 725-735. DOI: 10.1109/78.984771.

[19] M. Cevri., D. Üstündag, Bayesian Estimation of Sinusoidal Signals via Parallel Tempering, *8th Wseas International Conference on Signal Processing, Istanbul, Turkey*, 2009, pp. 67-72. DOI: 10.5555/1561963.1561974.

[20] D. Üstündağ, Recovering Sinusoids from Data Using Bayesian Inference with RJMCMC**,** *2011 Seventh International Conference on Natural Computation Shanghai, China*, 2011, pp.1-5. DOI: 10.1109/ICNC.2011.6022566

[21] P.M. Djurić, H.-T. Li, Bayesian Spectrum Estimation of Harmonic Signals, *IEEE Signal Processing Letters,* Vol. 2, 1995, pp. 213-215. DOI: 10.1109/97.473649.

[22] S.P. Brooks, A. Gelman, General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics,* Vol. 7, 4, 1998, pp. 434-455. DOI: 10.1080/10618600.1998.10474787.

[23] S. Suparman, Reversible Jump MCMC Method for Hierarchical Bayesian Model Selection in Moving Average model*, International Journal of Geomate*, Vol. 16, 56, 2019, pp. 9-15. DOI: 10.21660/2019.56.4509.

[24] H. Jeffreys, Theory of Probability, *Oxford University Press*, London, 1988. DOI: 10.1007/978-1-84800-048-3.

[25] J.M. Bernardo and A.F.M. Smith, Bayesian Theory, *Wiley series in Applied Probability and Statistics*, Wiley, New York, 1994. DOI: 10.1002/9780470316870.

[26] A.F. M. Smith, G.O. Roberts, Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of Royal Statistical Society B*, Vol. 35, 1, 1993, pp. 3-23. DOI: 10.1111/j.25176161.1993.tb01466.x.

[27] J. Vermaak, C. Andrieu, A. Doucet and S.J. Godsill, Reversible Jump Markov Chain Monte Carlo Strategies for Bayesian Model Selection in Autoregressive Processes, *Journal of Time Series Analysis*, Vol. 25, 2004, pp. 785-809. DOI: 10.1111/j.1467-9892.2004.00380.x.

[28] J.K. Joseph, O. Ruanaidh, W.J. Fitzgerald, Numerical Bayesian Methods Applied to Signal Processing, *Springer-Verlag*, New York, 1996. DOI: 10.1007/978-1-4612-0717-7.

[29] F. Goodyer, B.I. Ahmad and S. Godsill, Reversible Jump Markov Chain Monte Carlo for pulse fitting, *2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2024)*. Seoul, Korea DOI: 10.1109/ICASSP48485.2024.10448493.

[30] G. O. Roberts and J. S. Rosenthal, Optimal Scaling of Discrete Approximations to Langevin Diffusions, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 60, 1, 1998, pp. 255-268. DOI: 10.1111/1467-9868.00123.

[31] S.A. Sisson, Trans-Dimensional Markov Chains: A Decade of Progress and Future Perspectives, *Journal of the American Statistical Association*, Vol. 100, 2005, pp. 1077-1089. DOI: 10.2307/27590636.

[32] O. Karakus, E.E. Kuruoğlu, and M.A. Altınkaya, Beyond Trans-Dimensional RJMCMC With A Case Study İn İmpulsive Data Modeling, *Signal Processing*, Vol. 153, 2018, pp. 396-410. DOI: 10.1016/j.sigpro.2018.07.028.

[33] P.H. Baxendale, Renewal Theory and Computable Convergence Rates for Geometrically Ergodic Markov Chains, *Annals of Applied Probability*, Vol. 15, 1 B, 2005, pp.700-738. DOI: 10.1214/105051604000000710.

[34] H.T. Li, P. M. Djurić, An Iterative MMSE Procedure for Parameter Estimation of Damped Sinusoidal Signals, *Signal Processing*, Vol. 51, 1996, pp. 105-120. DOI: 10.1016/0165-1684(96)00035-7.

[35] C.L. Dominguez, V. J. Ramos, H.A. Rivera, E.P. Campos, J.M. Minjares, Y. Shmaliy, Denoising ECG Signals using Weighted Iterative UFIR Filtering, *WSEAS Transactions on Signal Processing,* Vol. 19, 2023, pp.148-157. DOI: 10.37394/232014.2023.19.16.

[36] B. Sarah, B. A. E. Hassen, Adaptive Gaussian Distribution Threshold Spatial Frequency Denoising Image, *WSEAS Transactions on Signal Processing,* Vol.17, 2021, pp.108-113. DOI: 10.37394/232014.2021.17.15.

[37] P. Wellin, R. Gaylord, and S. Kamin, *An Introduction to Programming with Mathematica,* Cambridge University Press, 2005. DOI: 10.1017/CBO9780511801303.

[38] S. Menendez, F. George, and B. M. G. Kibria, Performance of Some Modified Test Statistics for Testing the Population Signal to Noise Ratio: Simulation and Applications, *WSEAS Transactions on Signal Processing,* Vol.19, 2023, pp.83-102. DOI: 10.37394/232014.2023.19.10.

[39] J.W. Cooley, J.W. Tukey, An Algorithm for the Machine Calculation of Complex Fourier Series, *Mathematics of Computation, Vol. 19,*

90, 1965, pp. 297-301. DOI: 10.1090/S0025-5718-1965-0178586-1.

[40] J. Ye, A. Wallace and J. Thompson, Parallel Markov Chain Monte Carlo computation for varying-dimension signal analysis, 2009 17th European Signal Processing Conference, Glasgow, Scotland, 2009, pp. 2673-2677. DOI: 10.5281/ZENODO.41499.

[41] G.L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Lecture Notes in Statistics, vol. 48, Springer-Verlag, New-York, 1988. ISBN 0-387-96871-7.

[42] C. Rudahunga, H. Kiragu, M. Ahuna, Fusion Based MR Images Denoising Technique Using Frequency Domain and Non-Local Means Filters, *WSEAS Transactions on Signal Processing,* Vol. 18, 2022, pp.153-163. DOI: 10.37394/232014.2022.18.22.

APPENDIX

Table 1. The RJMCMC-SA algorithm

1.Initialization: Choose an initial temperature T_0 and some initial starting configuration
$k^{(0)}, \boldsymbol{\omega}_k^{(0)}, \delta^{2(0)}, \Lambda^{(0)}$.

2.Iteration i.

•SA

•RJMCMC step

•Sample $u: U(0,1)$

•If $u < u_k$

Update: update the radial frequencies $\boldsymbol{\omega}_k$ without altering the k number of components using diffffferent puposals in Section 4.1.3

$u < b_k$

Birth: add a sinusoid at random in Section 4.1.1

$u < d_k$

Death: remove a sinusoid at random in Section 4.1.1

$u < s_{k^i}$

Split: choose a sinusoid randomly and split it into two close sinusoids in Section 4.1.2

$u < m_{k^i}$

Merge: choose two close sinusoids randomly and merge them into one sinusoid in Section 4.1.2

3.Demarginalization: Updating noise variance σ_k^2 and amplitudes $\mathbf{a}_k$

4.Perform an SA **step** with Simulated annealing acceptance ratio in Section 4.2

•**Convergence**: Test the convergence in Section 5

•**Cooling:** Control the cooling schedule in Section 5

5. $i \leftarrow i+1$ and go to 2

Table 2. Output of Computer simulation of noisy two sinusoids

```
-----Bayesian Inference RJMCMC-SA-----
Standard Deviation of noise: Known
CPU:14.0156
Number of Observations: 64
Number of Parameters: 6
Estimated sTn ratio: 4.37008
Estimated sdv of noise (σ): 1.
```

Parameters	Estimated Values
ω[1]	1.25257 ± 0.00286756
ω[2]	1.357 ± 0.00327138
a[1]	5.04156 ± 0.210653
a[2]	4.17479 ± 0.210642

Table 3. Computer output of three noisy sinusoids

-----Bayesian Inference RJMCMC-SA---

Standard Deviation of noise: Known

CPU:16.0938

Number of Observations: 64

Number of Parameters: 9

Estimated sTn ratio: 20.4382

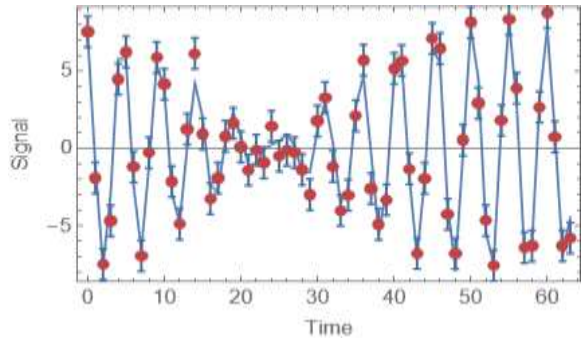
Estimated sdv of noise (σ): 1.

Parameters	Estimated Values
$\omega[1]$	1.25763 ± 0.0014001
$\omega[2]$	1.33287 ± 0.00904098
$\omega[3]$	1.45648 ± 0.00138092
a[1]	20.9351 ± 0.219515
a[2]	6.11828 ± 0.22325
a[3]	19.2027 ± 0.214409

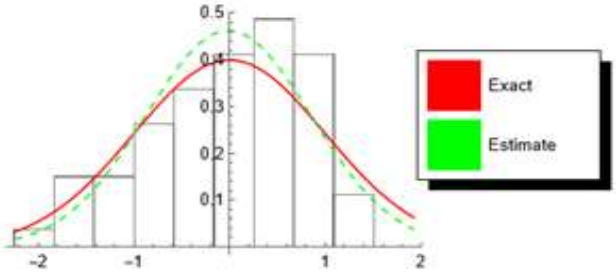
Table 4. Computer output of multiple noisy sinusoids

CPU: 1274.52s.
Number of Observations: 512
Number of Parameters: 15
STN ratio: 4.34691 $\hat{\sigma}$: 1

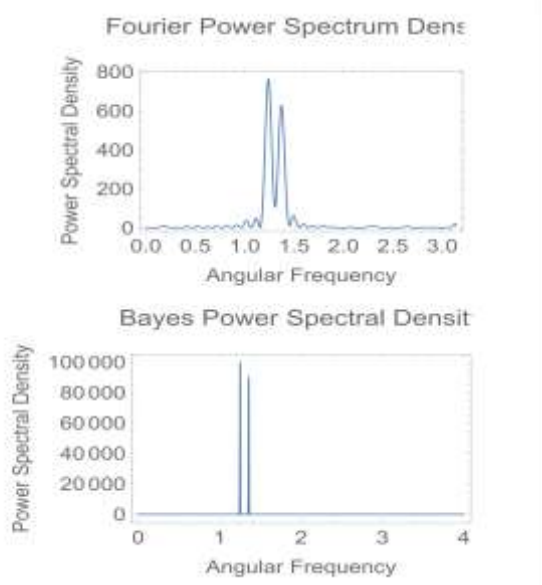
Parameters	True Values	Estimated Values
$\omega[1]$	0.1	0.0997618 ± 0.000519484
$\omega[2]$	0.15	0.149736 ± 0.000218277
$\omega[3]$	0.3	0.300139 ± 0.000233647
$\omega[4]$	0.31	0.310003 ± 0.000555051
$\omega[5]$	1	1.0002 ± 0.000137031
a[1]	1	0.839329 ± 0.0743751
a[2]	2	1.98232 ± 0.0743707
a[3]	5	4.93198 ± 0.0763934
a[4]	2	2.10663 ± 0.0763757
a[5]	3	3.10648 ± 0.0743281



(a)

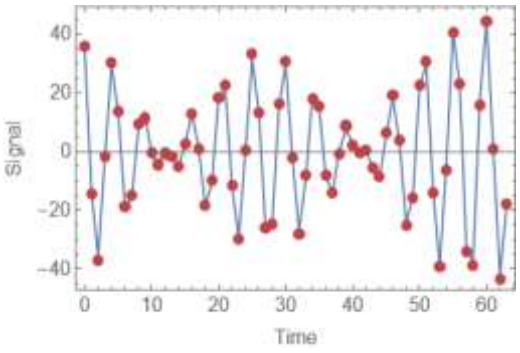


(b)

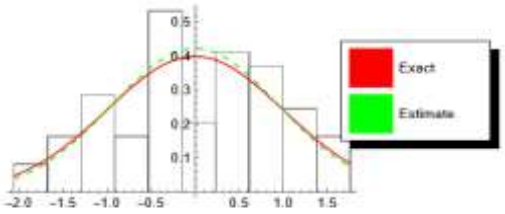


(c)

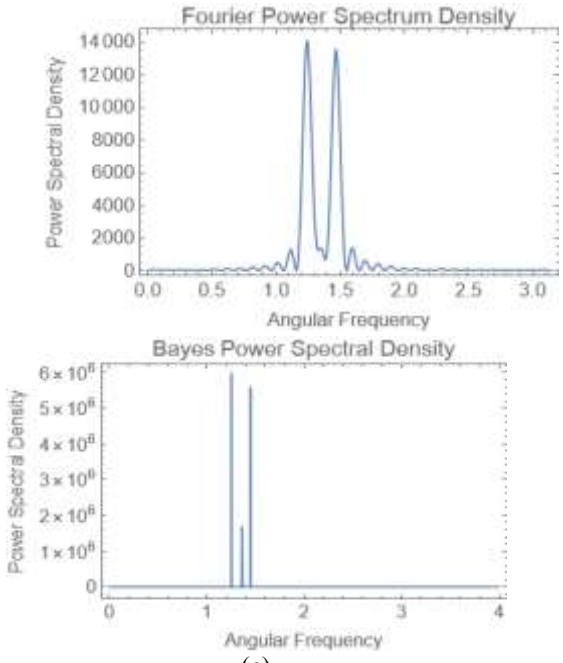
Fig. 1: (a) Red points show data samples from noisy two sinusoids. (b) Comparison of noise separate from data with its exact distribution. (c) Power spectral density varies with angular frequency



(a)



(b)



(c)

Fig. 2: (a) Red points show data samples from noisy three sinusoids. (b) Comparison of noise separated from data with its exact distribution. (c) Power spectral distribution varies with angular frequency

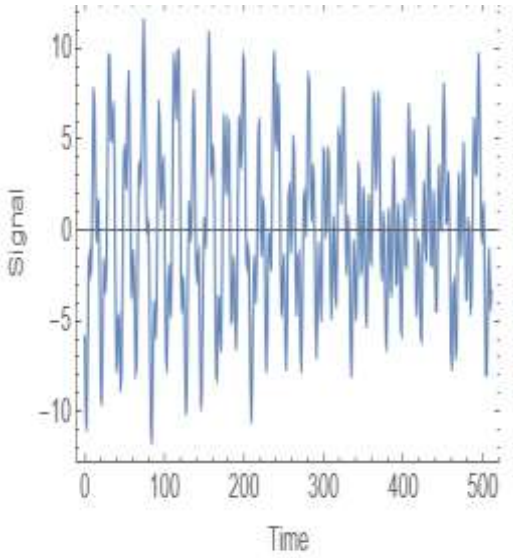
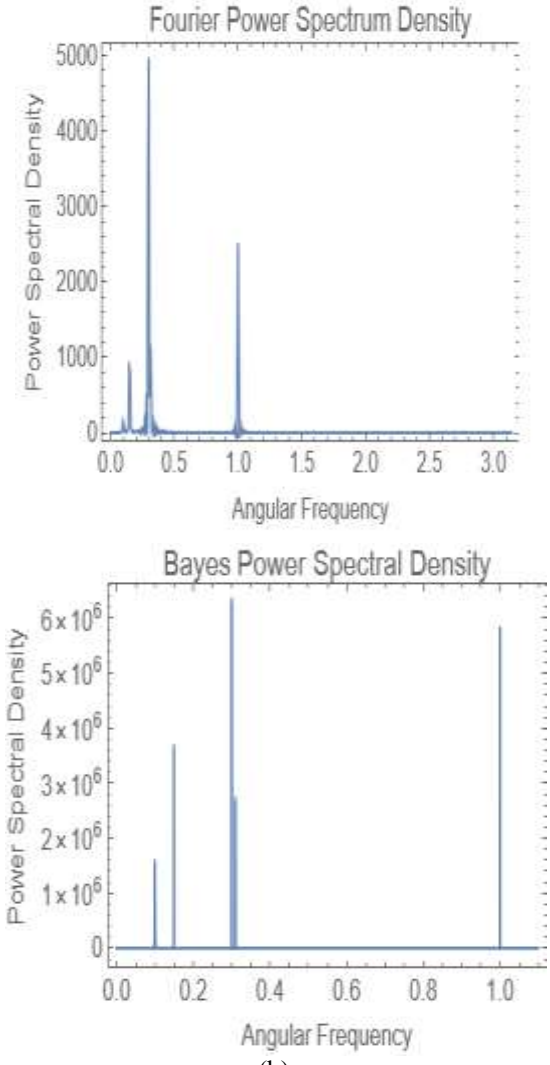


Fig. 3(a): Noisy data for multiple sinusoids



(b)

Fig. 3(b): Power spectral distribution varies with angular frequency for multiple sinusoids

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Dursun Üstündağ has contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.
- Mehmet Cevri has organized and executed the experiments of Section 5.

Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The authors have no conflicts of interest to declare.